

Облікова картка дисертації

I. Загальні відомості

Державний обліковий номер: 0826U004677

Особливі позначки: відкрита

Дата реєстрації: 18-08-2026

Статус: Наказ про видачу диплома

Реквізити наказу МОН / наказу закладу: UB КПІ ім. Ігоря Сікорського
№НСВС/66/26 від 21.09.2026 КЕП: Мельниченко А. А. 21.09.2026 11:36
3FAA9288358EC00304000000130730000B3AE100



II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Козирев Антон Юрійович

2. Anton Y. Kozyriev

Кваліфікація: 113

Ідентифікатор ORCID ID: 0009-0007-6692-2162

Вид дисертації: доктор філософії

Аспірантура/Докторантура: так

Шифр наукової спеціальності: 113

Назва наукової спеціальності: Прикладна математика

Галузь / галузі знань: математика та статистика

Освітньо-наукова програма зі спеціальності: Прикладна математика

Дата захисту: 31-08-2026

Спеціальність за освітою: Прикладна математика

Місце роботи здобувача: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, Київ, 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

III. Відомості про організацію, де відбувся захист

Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради): PhD 16673

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, Київ, 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

IV. Відомості про підприємство, установу, організацію, в якій було виконано дисертацію

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, Київ, 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

V. Відомості про дисертацію

Мова дисертації: Українська

Коди тематичних рубрик: 27.41.17, 27.47.19, 28.29.19, 28.29.23

Тема дисертації:

1. Методи стохастичної квантизації для масштабованого кластерного аналізу великих даних
2. Methods of Stochastic Quantization for Scalable Cluster Analysis of Big Data

Реферат/анотація:

1. Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 113 «Прикладна математика». – Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ. – 2026. Мета роботи – підвищення якості кластеризації та зменшення обчислювальних витрат під час оброблення великих даних, а також забезпечення стійкого до розподільного дрейфу ітеративного оновлення індексних структур у векторних базах даних без потреби в їх періодичній повній перебудові. Для досягнення мети в роботі визначено наступні наукові завдання: • аналіз існуючих методів кластеризації та виявлення їх переваг і недоліків при вирішенні задач кластерного аналізу; • розробка методу стохастичної квазіградієнтної кластеризації на основі транспортної задачі з рухомими центрами за допомогою стохастичної апроксимації Кіфера-Вольфовіца; • виведення умов збіжності запропонованого методу та дослідження швидкості збіжності з використанням елементів негладкого аналізу; • запровадження

структури даних стохастичного інвертованого файлового індексу на основі існуючої структури даних з використанням запропонованого методу для розбиття метричного простору вхідних даних; • проведення експериментальних досліджень на відкритих наборах даних Iris, MNIST та ImageNet з оцінкою точності за допомогою статистичних метрик, таких як індекс Ранда та середньоквадратична помилка. Наукова новизна. У дисертації одержано нові наукові результати: 1. Задачу кластеризації зведено до нової неопуклої негладкої задачі стохастичної оптимізації, що дозволяє для її розв'язання застосувати сучасні адаптивні методи стохастичної оптимізації, які використовуються для навчання глибоких нейронних мереж; 2. Розроблено новий метод стохастичної квазіградієнтної кластеризації, який використовує субдиференційовану цільову функцію транспортної відстані та стохастичну апроксимацію, що дозволяє оцінювати оптимальні розміщення центроїдів лише з використанням підмножини навчальних даних і потребує меншу кількість обчислювальних ресурсів. Запропонований метод має теоретично доведені асимптотичні гарантії збіжності на основі властивостей узагальненої диференційованості цільової функції; 3. Враховуючи, що метод стохастичної квазіградієнтної кластеризації має доведену збіжність лише до локальних екстремумів цільової функції, запропоновано новий метод початкової ініціалізації центроїдів за допомогою жадібного алгоритму в комбінації з дискретною оптимізацією та модифікацію стохастичної квазіградієнтної кластеризації з використанням метаевристичного методу диференціальної еволюції, що дозволяє здійснювати пошук глобальних екстремумів негладкої неопуклої цільової функції та підвищити якість кластеризації за рахунок уникнення збіжності до неоптимальних локальних розв'язків; 4. Удосконалено скінченно-різницевий метод оптимізації негладких неопуклих цільових функцій за допомогою методу згладжування (усереднення) функцій для пошуку точок екстремумів, використовуючи скінченно-різницеву апроксимацію градієнтів вздовж стохастичних напрямків. Теоретично доведено сублінійну швидкість збіжності стохастичного скінченно-різницевого методу для Ліпшицевих опуклих функцій; 5. Удосконалено структуру стохастичного інвертованого файлового індексу для інкрементного індексування векторних даних, де для індексації використовується запропонований метод стохастичної квазіградієнтної кластеризації, в умовах динамічної зміни відповідного розподілу даних через сезонні зміни, соціологічні фактори, тощо. Проблематика. Кластерний аналіз є основою сучасних систем штучного інтелекту, зокрема архітектур з пошуком за схожістю (RAG, RETRO) та векторних баз даних мільярдного масштабу, де інвертовані файлові індекси будуються переважно методом -середніх. Існуючі методи кластеризації на основі центроїдів мають низку суттєвих обмежень. По-перше, класичні методи -середніх, -медіан, гармонічних та нечітких -середніх потребують обчислювальної складності та пам'яті порядку на ітерацію, оскільки оновлення центроїдів вимагає повного проходження по всьому набору з векторів розмірності, що стає важкою задачею на масштабах порядку - на типовому обчислювальному обладнанні. По-друге, цільова функція внутрішньокластерної суми квадратів є неопуклою та задача пошуку її глобального мінімуму NP-важка, а класичні детерміновані методи збігаються лише до локального оптимуму, який сильно залежить від початкової ініціалізації центроїдів. Навіть удосконалення як -середніх++ забезпечують лише гарантію. По-третє, після побудови індексу його центроїди та межі розбиття залишаються незмінними, що в умовах розподільного дрейфу призводить до зростання помилки квантизації, дисбалансу розмірів кластерів і деградації якості пошуку.

2. Thesis for the degree of Doctor of Philosophy in the specialty 113 «Applied Mathematics». – National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv. – 2026. The purpose of the work is to improve the quality of clustering and reduce computational costs when processing big data, as well as to ensure iterative update of index structures in vector databases that is robust to distributional drift without the need for their periodic complete reconstruction. To achieve the stated purpose, the following research tasks have been defined: • conduct an analysis of existing clustering methods and identify their advantages and disadvantages in solving cluster analysis problems; • develop the stochastic quasigradient clustering method based on the transportation problem with moving centers using the Kiefer–Wolfowitz stochastic approximation; • derive convergence conditions of the proposed method and investigate the convergence rate using elements of non-smooth analysis; • introduce a stochastic inverted file index data structure based on an existing data structure using the proposed method to partition the metric space of the input data; • conduct experimental studies on open

datasets Iris, MNIST, and ImageNet and assess the accuracy of the obtained results using statistical metrics such as the Rand index and mean squared error. Scientific novelty. The thesis presents the following novel scientific results: 1. The clustering problem is reduced to a new non-convex non-smooth stochastic optimization problem, which allows for its solution to apply modern adaptive stochastic optimization methods used for training deep neural networks; 2. A new stochastic quasigradient clustering method is developed, which uses a subdifferentiated transport distance objective function and stochastic approximation, which allows estimating optimal centroid placements using only a subset of the training data and requires fewer computational resources. The proposed method has theoretically proven asymptotic convergence guarantees based on the properties of generalized differentiability of the objective function; 3. Considering that the stochastic quasi-gradient clustering method has proven convergence only to local extrema of the objective function, a new method of initial centroid initialization using a greedy algorithm in combination with discrete optimization and a modification of stochastic quasigradient clustering using the metaheuristic method of differential evolution is proposed, which allows searching for global extrema of a non-smooth non-convex objective function and improving the quality of clustering by avoiding convergence to suboptimal local solutions; 4. The finite-difference method of optimizing non-smooth non-convex objective functions is improved using the method of smoothing (averaging) functions to search for extrema points, using finite-difference approximation of gradients along stochastic directions. The sublinear convergence rate of the stochastic finite-difference method for Lipschitz convex functions has been theoretically proven; 5. The structure of the stochastic inverted file index for incremental indexing of vector data has been improved, where the proposed stochastic quasi-gradient clustering method is used for indexing, under conditions of dynamic change of the corresponding data distribution due to seasonal changes, sociological factors, etc. Problem statement. Cluster analysis constitutes the foundation of modern artificial intelligence systems, particularly similarity-search architectures (RAG, RETRO) and billion-scale vector databases, where inverted file indices are constructed predominantly by means of the k -means method. Existing centroid-based clustering methods exhibit a number of substantial limitations. First, the classical k -means, k -medians, harmonic and fuzzy k -means methods exhibit per-iteration computational and memory complexity of order $O(n^2)$, since updating the centroids requires a full pass over the entire set of vectors of dimension n , which becomes infeasible on commodity hardware as n reaches 10^6 . Second, the within-cluster sum-of-squares objective is non-convex, finding its global optimum is NP-hard, and deterministic methods converge only to a local optimum whose quality depends strongly on the initial choice of centroids; even refinements such as k -means++ provide only an guarantee. Third, once an stochastic inverted file index has been built, its centroids and partition boundaries remain frozen, so that under dynamic distributional drift (seasonal changes, sociological factors, evolving content) the quantization error grows, the partition sizes become unbalanced, and the quality of downstream retrieval and generative models deteriorates.

Державний реєстраційний номер ДіР:

Пріоритетний напрям розвитку науки і техніки: Інформаційні та комунікаційні технології

Стратегічний пріоритетний напрям інноваційної діяльності: Розвиток сучасних інформаційних, комунікаційних технологій, робототехніки

Підсумки дослідження: Теоретичне узагальнення і вирішення важливої наукової проблеми

Публікації:

- Norkin, V., Pichler, A., & Kozyriev, A. (2025). Constrained Global Optimization by smoothing. Lecture Notes in Computer Science, Vol. 14476. P.136–150. Print ISSN 0302-9743.
- Norkin, V., Kozyriev, A. and Norkin, B. (2024) “Modern stochastic quasi-gradient optimization algorithms”, International Scientific Technical Journal "Problems of Control and Informatics", 69(2), pp. 71–83.
- Kozyriev, A., & Norkin, V. (2024). Robust Clustering on High-Dimensional Data with Stochastic Quantization. International Scientific Technical Journal “Problems of Control and Informatics,” 70(1), pp. 32–48.

- Kozyriev, A., & Norkin, V. (2024). Lossy image compression with stochastic quantization. Cybernetics and Computer Technologies, (3), 60–66.
- Norkin, V., & Kozyriev, A. (2023). On Shor's R-algorithm for problems with constraints. Cybernetics and Computer Technologies, (3), 16–22.

Наукова (науково-технічна) продукція: методи, теорії, гіпотези; програмні продукти, програмно-технологічна документація; аналітичні матеріали

Соціально-економічна спрямованість: підвищення автоматизації виробничих процесів; забезпечення промисловості чи населення новим видом інформаційно-комунікаційних послуг; підвищення ефективності систем підтримки прийняття рішень

Охоронні документи на ОПІВ:

Комп'ютерні програми

Програмне забезпечення методу стохастичної квазі-градієнтної кластеризації, його модифікацій з адаптивним кроком та моделі структури даних стохастичного інвертованого файлового індексу

Наукові відкриття

Математична модель стохастичної квазі-градієнтної кластеризації

Впровадження результатів дисертації: Впроваджено

Зв'язок з науковими темами:

VI. Відомості про наукового керівника/керівників (консультанта)

Власне Прізвище Ім'я По-батькові:

1. Норкін Володимир Іванович

2. Volodymyr I. Norkin

Кваліфікація: д. ф.-м. н., старший науковий співробітник, 01.05.01

Ідентифікатор ORCID ID: 0000-0003-3255-0405

Додаткова інформація:

Повне найменування юридичної особи: Інститут кібернетики імені В. М. Глушкова Національної академії наук України

Код за ЄДРПОУ: 05417176

Місцезнаходження: проспект Академіка Глушкова, Київ, 03187, Україна

Форма власності:

Сфера управління: Національна академія наук України

Ідентифікатор ROR: Не застосовується

VII. Відомості про офіційних опонентів та рецензентів

Офіційні опоненти

Власне Прізвище Ім'я По-батькові:

1. Малик Ігор Володимирович

2. Ihor V. Malyk

Кваліфікація: д. ф.-м. н., професор, 01.05.01

Ідентифікатор ORCID ID: 0000-0002-1291-9167

Додаткова інформація:

Повне найменування юридичної особи: Чернівецький національний університет імені Юрія Федьковича

Код за ЄДРПОУ: 02071240

Місцезнаходження: вул. Коцюбинського, Чернівці, 58012, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Власне Прізвище Ім'я По-батькові:

1. Стовба Віктор Олександрович

2. Viktor O. Stovba

Кваліфікація: д.філософ, 113

Ідентифікатор ORCID ID: 0000-0003-3023-5815

Додаткова інформація: <https://www.scopus.com/authid/detail.uri?authorId=57205705471>

Повне найменування юридичної особи: Інститут кібернетики імені В. М. Глушкова Національної академії наук України

Код за ЄДРПОУ: 05417176

Місцезнаходження: проспект Академіка Глушкова, Київ, 03187, Україна

Форма власності:

Сфера управління: Національна академія наук України

Ідентифікатор ROR: Не застосовується

Рецензенти

Власне Прізвище Ім'я По-батькові:

1. Пишнограєв Іван Олександрович

2. Ivan O. Pyshnohraiev

Кваліфікація: к. ф.-м. н., доц., 01.05.04

Ідентифікатор ORCID ID: 0000-0002-3346-8318

Додаткова інформація: <https://scholar.google.com/citations?user=9qVIOVYAAAAJ&hl=en>

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, Київ, 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

Власне Прізвище Ім'я По-батькові:

1. Хайдуров Владислав Володимирович

2. Vladyslav V. Khaidurov

Кваліфікація: к. т. н., с.д., 01.05.02

Ідентифікатор ORCID ID: 0000-0002-4805-8880

Додаткова інформація:

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Берестейський, Київ, 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

VIII. Заключні відомості

**Власне Прізвище Ім'я По-батькові
голови ради**

Чертов Олег Романович

**Власне Прізвище Ім'я По-батькові
головуючого на засіданні**

Чертов Олег Романович

**Відповідальний за підготовку
облікових документів**

Козирев Антон Юрійович

Реєстратор

Юрченко Тетяна Анатоліївна

**Керівник відділу УкрІНТЕІ, що є
відповідальним за реєстрацію наукової
діяльності**



Юрченко Тетяна Анатоліївна